

# TPS: A Topological Potential Scheme to Predict Influential Network Nodes for Intelligent Communication in Social Networks

Yuxin Mao<sup>ID</sup>, Lujie Zhou, and Naixue Xiong<sup>ID</sup>

**Abstract**—The growing popularity of Online Social Networks (OSN) have prompted an increasing number of companies to promote their brands and products through social media. This paper presents a topological potential scheme for predicting influential nodes from large scale OSNs to support more intelligent brand communication. We first construct a weighted network model for the users and their relationships extracted from the brand-related content in OSNs. We quantitatively measure the individual value of the nodes from both the network structure and brand engagement aspects. Moreover, we have addressed the problem of influence decay along with information propagation in social networks and use the topological potential theory to evaluate the importance of the nodes by their individual values as well as the individual values of their surrounding nodes. The experimental results have shown that the proposed method is able to predict influential nodes in large-scale OSNs. We investigate the top-k influential nodes identified by our method in detail, which are quite different from those identified by using pure network structure or individual value. We can obtain an identification result with a higher ratio of verified users and user coverage by using our method compared to existing typical approaches.

**Index Terms**—Social networking, data mining, topology, intelligent communication, influential network nodes.

## I. INTRODUCTION

ONLINE Social Networks (OSN) have become increasingly popular in recent years. With the emergence of the mobile Internet, users are able to enjoy OSNs such as Facebook, Twitter, and Weibo at all times and in all places. Extensive online User-Generated Content (UGC) has been produced on social media, and has become an important current aspect of Electronic Word of Mouth (eWOM). Social media has become an important channel through which companies can release information to and maintain contact with their customers. Therefore, eWOM via social media has become a key driver of brand marketing towards consumers,

Manuscript received October 8, 2020; revised November 28, 2020; accepted December 7, 2020. Date of publication December 14, 2020; date of current version March 17, 2021. This work was supported in part by the National Social Science Foundation of China under Grant 16BGL193 Recommended for acceptance by Dr. Pan Zhou. (*Corresponding author: Naixue Xiong.*)

Yuxin Mao and Lujie Zhou are with the School of Management and E-Business, Zhejiang Gongshang University, Hangzhou 310018, China (e-mail: maoyuxin@zjgsu.edu.cn; jasminezlj@qq.com).

Naixue Xiong is with the Department of Mathematics and Computer Science, Northeastern State University, Tahlequah, OK 74464 USA (e-mail: xionгнаixue@gmail.com).

Digital Object Identifier 10.1109/TNSE.2020.3044299

prompting an increasing number of companies to promote their brands and products through OSNs. From the marketing perspective, the importance of the nodes in a large-scale OSN is not equal. There exist some active users in the network, who have a certain influence and are also very concerned about some brands. Obviously, these influential nodes can help companies to perform brand communication through social media by affecting other nodes. Therefore, if the influential nodes can be identified within large-scale OSNs, then companies can rely on them for brand communication. Those influential nodes will act as ‘bridges’ between companies and other consumers.

Although there have been a number of previous studies about identifying or predicting influential nodes in OSNs [1], [2], few have addressed the potential significance to brand communication or how to identify influential nodes that are more suitable for promoting brands through social media. Moreover, massive brand-related data have become a kind of big data [3] in OSNs. Therefore, predicting influential nodes within a large-scale OSN for brand communication is still a problem worthy of further study. In this paper, we propose a topological potential scheme for predicting influential nodes in OSNs by considering both the network structure and brand engagement factors. The preliminary results of this study can help companies analyze and discover the characteristics and rules of OSNs to provide decision support for data-driven or intelligent brand communication in social media. In particular, we have considered the problem of influence decay in OSN and apply a topological potential model to identify influential nodes more suitable for brand communication. The major contributions of this study are summarized as follows:

- 1) We propose to measure the importance of nodes in OSN by considering both the network structural and content-related metrics and quantitatively represent it as individual value.
- 2) An intelligent topological potential scheme (TPS) is proposed to determine the node influence and predict influential nodes in OSNs for brand communication.
- 3) We collected a real-world dataset from SMZDM.com including more than 40000 users and 60000 social relations. Comprehensive experiments are conducted to validate the effectiveness of our method.

The rest of the paper is organized as follows. In Section II, the motivations are introduced, and the related works are reviewed. Section III describes the process of predicting

influential nodes in OSN for brand communication in detail. Details about the performance evaluation are presented in Section IV. Finally, some conclusion and future work are presented in Section V.

## II. RELATED WORKS

Currently, many efforts have been made to identify or predict influential nodes in OSNs. In this section, we briefly review the existing works in several categories.

### A. Structural Methods

Social network analysis mostly relies on topological metrics [4] such as centrality and community concepts, and many of the terms used to measure these metrics are a reflection of their sociological origin [5]. For example, Freeman [6], [7] illustrates that the centrality of a node indicates the connection ability of the node in the social network structure and can be used as a criterion for measuring the importance of the node. Corley & Sha [8] address the problem of  $n$ -most vital nodes problem and propose the algorithm to solve the problem of node importance evaluation. Currently, many efforts have been made to discover the most influential nodes for maximizing influence in social networks [9]–[11]. These studies of influence maximization aim to discover nodes that can activate as many nodes as possible, which indicates that the influence of nodes can be propagated as extensively as possible.

For example, Zareie et al. [12] introduce two influential node ranking algorithms that use the diversity of the neighbors of each node [13] to obtain its ranking value. Kumar & Panda [14] propose a coreness-based method to find influential nodes by voting. They also compare the performance of their method with some existing popular methods. Salavati & Abdollahpouri [15] take into account the interactions between users and network topology in weighted and directed graphs and consider target users' profit and similarity in identifying influential nodes. Zhang et al. [16] introduce a trust-based influential node discovery method for identifying influential nodes in social networks. However, their idea about trust between nodes is still based on the topological information of the network. Salavaty et al. [17] develop a formula that integrates the most significant network centrality measures in order to synergize their effects and simultaneously remove their biases to identify the most influential nodes in a complex network. However, their method is mainly used in biological systems. Zhou et al. [18] intended to solve the problem of finding the influential nodes which are able to initiate large-scale spreading processes in a limited amount of time. Amnieh and Kaedi [19] try to use two personality characteristics, openness and extroversion, to estimate for network members and find influential nodes. However, their personality characteristics are still computed based on the network structure.

There are also a few of methods that take into account the influence of community (or group) structure [20], [21] in the network. Jain & Katarya [22] identify the community structure within the social network and the opinion leader by using a modified firefly algorithm in each community. Srinivas & Rajendran [23] propose an integer linear programming model to detect community structure in real-life networks and also

identify the most influential node within each community. Zhao et al. [24] propose an algorithm for identifying influential nodes in social networks with community structure based on label propagation. The proposed algorithm can find the core nodes of different communities in the network through the label propagation process. Generally, these methods identify global influential users regardless of domain-specific information.

### B. Hybrid Methods

The spreading influence of a node on a network depends on a number of factors, including its location on the network, the content of exchanged messages [25], and the character and amount of activity of the node [12]. Therefore, pure network structural methods are quite insufficient for identifying influential nodes in OSNs. In contrast, hybrid methods combining network structure and content seem to be more suitable for this problem. For example, Aleahmad et al. [1] try to detect the main topics of discussion in a given domain, calculate a score for each user, and then calculate a probability of being an opinion leader by using the scores. Liu et al. [26] take into account the dimensions of trust, domain, and time, and propose a product review domain-aware approach to identify effective influencers in OSNs. Advertising cost has also been taken into account, in addition to nodes influentiality, to determine influential users [27], [28]. Zareie et al. [2] measure the interest of users in marketing messages and then propose an algorithm to obtain the set of the most influential users in social networks. Weng et al. [29] propose an extension of PageRank algorithm called TwitterRank, to measure the influence of users in Twitter. They measure the influence taking both the topical similarity between users and the link structure into account.

Moreover, many researchers have tried to use the ranking model like PageRank to identify opinion leader detection and especially in combination with topic models, e.g., Influence Rank [30], OpinionRank [31], Dynamic OpinionRank [32], TopicSimilarRank [33] and others. SuperedgeRank [34] is a mixed framework to find the influential users based on super-network theory, that is composed of network topology analysis and text mining. Li et al. [35] develop a ranking framework to automatically identify topic-specific opinion leaders. The score for opinion leadership is computed from four measures include expertise, novelty, influence, and activity. Topic-based methods can also be used to mine influential users in OSNs. For example, Hamzehei et al. [36] propose a topic-based influence measurement approach to integrate the user-topic relationships, topic content information, and social connections between users. Fang et al. [37] address the more important topic-level influence and develop a topic-sensitive influencer mining framework in interest-based OSNs.

Although these hybrid methods may gain better performance by combining network structural features and content-related features, most of them haven't addressed the problem of influence decay along with information propagation [38]. In other words, we should consider the influence of users from the perspective of dynamics in information propagation, rather than single and static user.

### C. Brand Marketing in Social Media

In addition to the methods mentioned above, many efforts have been made to study how social media can be used to support brand communication or how brands can be promoted in social media in the field of marketing. For example, Hajikhani *et al.* [39] try to investigate the overall polarity of public sentiment regarding specific companies' products by analyzing content from Twitter. Kabadayi & Price [40] study the factors affecting consumers' liking and commenting behaviors on Facebook brand pages. Schivinski & Dabrowski [41] investigate 504 Facebook users in order to observe the impact of firm-created and user-generated social media communication on brand equity, brand attitude and purchase intention by using a standardized online survey. Jiménez-Castillo & Sánchez-Fernández [42] study how effective digital influencers are in recommending brands via electronic word-of-mouth by examining whether the potential influence they have on their followers may affect brand engagement. Gao & Feng [43] examine the differences in Chinese users' gratifications of different social media and the impact of brand content strategies on the quality of brand-consumer communication via social media. Godey *et al.* [44] study how social media marketing activities influence brand equity creation and consumers' behavior towards a brand. Veirman *et al.* [45] explore the marketing through Instagram influencers and assess the impact of number of followers and product divergence on brand attitude by two experiments with fictitious influencer accounts on Instagram. Although many studies have been done about brand communication in social media, fewer existing studies have addressed how to identify and make use of influential nodes for brand marketing on social media.

## III. OUR PROPOSED TPS

In this section, we mainly present an intelligent method for predicting influential nodes in OSN for intelligent brand communication.

### A. Weighted Network Model

An OSN can be formally represented as a graph  $G = (V, E, W)$ , where  $V$  denotes the set of people or users that belong to the network and  $E$  represents the set of relations between the users. There is an edge between two nodes if they have a social relation. Given two nodes  $u_i$  and  $u_j$ , if  $u_j$  follows  $u_i$ , then there is an edge directed from  $u_i$  to  $u_j$ . Moreover, if the post of a user is commented on by another user, we consider this interaction as another kind of social relation between two users. For example, if  $u_j$  comments on a post generated by  $u_i$ , then there is an edge directed from  $u_i$  to  $u_j$ . If  $u_j$  follows  $u_i$  or comments on  $u_i$ 's post, it means that  $u_i$  is able to affect  $u_j$  or that information can spread from  $u_i$  to  $u_j$ .  $W$  indicates a set of weights for the directed edges in  $E$ . The value of the weights in  $W$  denotes the number of relations and interactions between the users.

For a specific brand (e.g., a cell phone or cosmetics), we can extract all posts related to it from an OSN and construct a corresponding weighted network model before we start to identify the influential nodes. Then, the task of mining influential nodes can be constrained in a limited space or community.

The detailed process of network model construction is illustrated as follows:

- 1) First, we crawl the posts about the brand within a period of time (e.g., one month) from an OSN and the set of posts are denoted as  $P$ .
- 2) Then we extract the authors of the posts in  $P$ , and get a set of users, so called  $U$ .
- 3) The relations between the users in  $U$  are further extracted and added to a set  $R$ . Each  $r$  in  $R$  can be denoted as  $\langle u_i, u_j \rangle$ , where  $u_i$  and  $u_j$  are the two users have a social relation  $r$  in the OSN. To each  $r$  in  $R$ , we create a corresponding weight  $w$ , set  $w = 1$  and add  $w$  to a weight set  $W$ .
- 4) To each user  $u$  in  $U$ , we get the users who follow  $u$  in the OSN as a set  $U_{uf}$ . We also get all his/her posts in  $P$ , marked as  $P_u$  and we have  $P_u \subset P$ . To each post  $p$  in  $P_u$ , we get all the users who have commented on  $p$  as a set  $U_{uc}$ . Then we have an extended user set  $U_u = U_{uf} \cup U_{uc}$  for  $u$ .
- 5) To each user  $u_i$  in  $U_u$ , if  $\langle u, u_i \rangle \in R$ , then we find the corresponding weight  $w_i$  for  $\langle u, u_i \rangle$  in  $W$ , perform  $w_i = w_i + 1$  and update the value of  $w_i$  in  $W$ . If  $\langle u, u_i \rangle \notin R$ , it means that this is a newly-found social relation. In this case, we add  $\langle u, u_i \rangle$  to  $R$ , create a new weight  $w_i = 1$  for  $\langle u, u_i \rangle$  and add  $w_i$  to  $W$ . Moreover, we add  $u_i$  to a temporary set  $U'$ .
- 6) We update the user set  $U$  by performing  $U \cup U'$ .
- 7) Finally, we get the weighted network model  $G = (U, R, W)$  for a specific brand in the OSN. Here  $U$  and  $R$  can also be represented by  $V$  and  $E$ .

### B. Network Structure Characteristics

In this article, we take into account two typical and frequently-used structural metrics to support our method, namely, *outdegree* and *betweenness centrality*. These two metrics can be used to measure the scope of nodes' influence and their ability to control the community in the network.

Given a network  $G = (V, E, W)$ , the outdegree of a node can be formally denoted by the following equation:

$$od(u_i) = \sum_{j \in N} r(u_i, u_j) w_{i,j}, \quad (1)$$

where  $u_i$  and  $u_j$  represent two nodes in the network,  $r(u_i, u_j) \in E$  represents a directed edge from  $u_i$  to  $u_j$ ,  $w_{i,j} \in W$  represents the weight of the edge, and  $N \subset V$  represents the adjacent node set of  $u_i$ .

The outdegree of a node is mainly related to the behaviors of following and commenting. Users can follow others whom they are interested in. To an active user  $u_i$ , the more other users who follow  $u_i$ , the more attractive  $u_i$  is and thus the greater ability he/she has to influence others. Users can also comment on the posts about which they are concerned. Given a post  $p_j$  generated by user  $u_i$ , the more comments that  $p_j$  gets, the wider the scope of influence of  $p_j$  is. The more times that  $u_i$ 's posts are commented on, the greater influence the information generated by  $u_i$  has.

Given three nodes  $u_i u_j u_k$ , then the control ability of  $u_i$  over the communication between  $u_j$  and  $u_k$  is computed by the following equation:

$$ca_{jk}(u_i) = \frac{g_{jk}(u_i)}{g_{jk}}, \quad (2)$$

where  $g_{jk}$  represents the total number of shortest paths between  $u_j$  and  $u_k$ , and  $g_{jk}(u_i)$  represents the total number of shortest paths between  $u_j$  and  $u_k$  passing through  $u_i$ . Note that we only consider the case that there exists at least one path between the two nodes  $u_j$  and  $u_k$ .

We can calculate the sum of the control capability of  $u_i$  with respect to all node pairs in the network and finally obtain the betweenness centrality of  $u_i$  as follows:

$$bc(u_i) = \sum_j \sum_k ca_{jk}(u_i), \quad j \neq i \neq k. \quad (3)$$

The betweenness centrality of a node considers the degree that counts the occurrence of a node on the straight (or shortest) path between other nodes. That is, if a node is the only way for other nodes in the network to connect with others, it has a more important position in the network. Given an active user  $u_i$ , the larger the betweenness centrality of  $u_i$  is, the more important location he/she has in the network.

As we have mentioned before, the weight of an edge represents the closeness of the relationship between the two nodes. To simplify the calculation of the distance between nodes, we first determine the maximum edge weight  $w_{max}$  in the original network and then use the following equation to update the original weight for each edge:

$$w'_i = w_{max} + 1 - w_i. \quad (4)$$

In this way, we obtain an updated weight set  $W'$  for the network. For any node pair  $u_i$  and  $u_j$  in the network, we use an improved Floyd algorithm to calculate all the shortest paths and the corresponding shortest distances between the two nodes. Then, we can calculate the betweenness centrality for each node.

To avoid the impact of excessive difference between the two metrics, we perform a maximum-minimum normalization on the two metrics as follows such that both metrics are mapped to the interval  $[0, 1]$ :

$$f_n = \frac{(f - f_{min})}{(f_{max} - f_{min})}. \quad (5)$$

Therefore, we can get the overall network structure score for a node  $u_i$  by the following equation:

$$score_{network}(u_i) = \frac{od_{norm}(u_i) + bc_{norm}(u_i)}{2}, \quad (6)$$

where  $od_{norm}(u_i)$  refers to the normalized value of outdegree and  $bc_{norm}(u_i)$  refers to the normalized value of betweenness centrality. A larger  $score_{network}(u_i)$  value implies that node  $u_i$  has a more important location in the network from the structural perspective.

### C. Brand Engagement-Based Value

In the context of brand communication, only considering the network structural metrics is insufficient to discover the real influential nodes. We should also take into account the content-related metrics to measure the individual value of the nodes in OSN. To identify influential nodes that are suitable for the communication and marketing of a specific brand, we should check whether a user is concerned about the brand. Therefore, we try to measure the value of nodes from the perspective of brand loyalty [46] or brand engagement [47] in addition to network structure. We try to quantitatively measure the brand engagement-based value of a node. As brand engagement is directly related to users' behaviors [48] in OSNs, we mainly consider the following four behaviors:

- 1) *Publishing*: A user writes or shares posts.
- 2) *Commenting*: A user comments on the posts by others.
- 3) *Liking*: A user presses the 'like' button below a post.
- 4) *Adding to favorites*: A user adds a post to his/her favorites.

It is not difficult to quantify the above behaviors. Given a brand  $b_j$  and a user  $u_i$ , we can obtain the number of posts related to  $b_j$  that  $u_i$  has actively published on his/her personal page. As a potential influential node, he/she shall publish and share information related to a certain brand (product, event, etc.) frequently. Moreover, we can also obtain the percentage of positive posts related to  $b_j$  published by  $u_i$ , which are positively commented on, liked and added to favorites by other users. If many users positively respond to the posts, it reflects that  $u_i$  is able to evoke the emotional resonance of other users or obtain their support for  $b_j$ . We illustrate how to measure brand engagement quantitatively by the following steps:

- 1) *Mark the polarity of posts*: If the post content is negative about the brand, we mark the post as negative or with '-'. Similarly, if the post content is nonnegative about the brand, we mark the post as nonnegative or with '+'.
  - 2) *Calculate the support rate of posts*: A semantic analysis approach based on sentiment dictionary is used to evaluate the opinions of other users on specific posts. We evaluate the sentiment polarity of each comment on a post and classify the sentiment polarity into negative and nonnegative. Then we calculate the support rate of posts ( $p_{support}$ ) by the following equation:

$$p_{support} = \frac{N_{pos\_com} + N_{favorite} + N_{like}}{N_{pos\_com} + N_{neg\_com} + N_{favorite} + N_{like}}, \quad (7)$$

where,  $N_{pos\_com}$  is the number of nonnegative comments,  $N_{neg\_com}$  represents the number of negative comments,  $N_{favorite}$  represents the number of adding to favorites, and  $N_{like}$  represents the number of likes.

- 1) *Obtain the brand engagement-based value for a user*: Then, we can obtain the overall brand engagement score for node  $u_i$  by using the following equation:

$$score_{brand} = \sum_i \left( post_{polar}^i \times p_{support}^i \right), \quad (8)$$

where,  $i$  represents the  $i$ -th brand-related post published by  $u_i$ ,  $post_{polar}^i$  represents the polarity of the  $i$ -th post, and  $p_{support}^i$  represents the support rate of the  $i$ -th post.

#### D. Measuring a Node's Individual Value

After evaluating each node's characteristics, we can obtain the individual value of each node by the weight sum of the scores of each factor. We can use entropy theory to determine the weight for the two scores of a node, the so-called entropy weight, and then make a comprehensive and objective evaluation of the individual value of the node.

Given  $n$  nodes in a network with two scores each, we can construct an  $n * 2$  matrix  $R$ . Each row in  $R$  represents a node, each column represents a score, and item  $r_{ij}$  in  $R$  represents the  $j$ -th influence value of the  $i$ -th node. Let  $f_{ij} = \frac{r_{ij}}{\sum_{i=1}^n r_{ij}}$  and  $\mu = \frac{1}{\ln n}$ , with  $f_{ij} = 0$  and  $f_{ij} \ln(f_{ij}) = 0$ . The entropy value of the  $j$ -th influence value is defined as follows:

$$H_j = -\mu \sum_{i=1}^n f_{ij} \ln(f_{ij}), (j = 1, 2). \quad (9)$$

Then, the entropy weight of the  $j$ -th influence value is defined as follows:

$$w_j = \frac{1 - H_j}{2 - \sum_{j=1}^2 H_j}, (j = 1, 2) \quad (10)$$

We can further measure the individual value of the node by the following equations:

$$\gamma_i = \sum_{j=1}^2 r_{ij} * w_j, \quad (11)$$

$$\Rightarrow value_{indv} = score_{network} * w_1 + score_{brand} * w_2. \quad (12)$$

As we can calculate the individual value for each node in the network, the individual value of the users can be represented as the weights of the nodes. Therefore, we can obtain a dual-weighted network model for brand communication. The corresponding formal representation for the dual-weighted model is as follows:

$$G' = (V, E, W', A),$$

where  $W'$  represents the updated weight set for  $E$  according to 4, and  $A = \{a_1, a_2, \dots, a_n\}$  represents the set of individual values for the nodes in  $V$ .

The ultimate purpose of mining influential nodes in big data is to support more intelligent brand marketing. Thus, influential nodes should have a stronger ability to disseminate marketing information for a brand. Although we have proposed to use individual value to measure the importance of each node in OSN, we still cannot guarantee that a node with high individual value always disseminate information efficiently. For example,  $u$  is a node with high individual value, but the individual values of the nodes around  $u$  are very low. In this case,

---

#### Algorithm I: Predicting Influential Nodes

---

**Input:** Network Model  $G = (V, E, W)$

**Output:** Top n% Influential Node Set  $R_{top}$

- 1: get the updated weight set  $W'$
  - 2: for each node pair  $u_i$  and  $u_j$  ( $i \neq j$ ) in  $V$
  - 3: calculate the shortest distance  $d_{ij}$  between  $u_i$  and  $u_j$  by the Floyd algorithm
  - 4: end for
  - 5: for each node  $u_i$  in  $V$
  - 6: calculate  $value_{indv}$  for  $u_i$ ;
  - 7: add  $value_{indv}$  to a set  $A$ ;
  - 8: end for
  - 9: get the updated network model  $G' = (V, E, W', A)$
  - 10: for each node  $u_i$  in  $V$
  - 11: calculate  $\Phi(u_i)$ ;
  - 12: if  $(\Phi(u_i) \geq \Phi_{threshold})$
  - 13: add  $u_i$  to  $R_{infl}$ ;
  - 14: end if
  - 15: end for
  - 16: sort the items in  $R_{infl}$  by  $\Phi$  value in descending order;
  - 17: get the first n% items in  $R_{infl}$  and add them to a new set  $R_{top}$ ;
  - 18: return  $R_{top}$ ;
- 

the marketing information originated from  $u$  may not spread well in the network, as the information dissemination capacity of its surrounding nodes is not strong enough. In other words, although the individual value of  $u$  is high, we still cannot consider it as an influential node due to the low individual values of its surrounding nodes. Therefore, when we determine whether a node is an influential node, we should also consider not only the individual value of the node but also the individual values of its surrounding nodes. Nodes with high individual values can obviously affect their surrounding nodes, but this effect will decay as the distance increases [49]. Therefore, we need more replay nodes with high individual values to support more efficient information spreading or dissemination in the network [50]–[52].

To address this issue, we try to further make use of topological potential theory to determine influential nodes in our method. According to the topological potential theory, a node will be affected by other nodes in the network. We improve the typical topological potential equation and calculate the topological potential value as follows:

$$\Phi(u_i) = \sum_{j=1}^n \left( v_i \times v_j \times e^{-\left(\frac{d_{ij}}{\sigma}\right)^2} \right), \quad (13)$$

where  $d_{ij}$  denotes the shortest distance between nodes  $u_i$  and  $u_j$ , influence factor  $\sigma$  is a parameter used to depict the influence range of each node;  $v_i$  refers to the individual value of node  $u_i$ ,  $v_j$  refers to the individual value of node  $u_j$ , and  $\Phi(u_i)$  is the topological potential value of  $u_i$ .

The potential entropy can be calculated as follows:

$$H = - \sum_{i=1}^n \frac{\varphi(v_i)}{Z} \ln \left( \frac{\varphi(v_i)}{Z} \right), \quad (14)$$

where  $Z = \sum_{i=1}^n \varphi(v_i)$  is a normalization factor.

If we put 13 into 14, the potential entropy  $H$  is a function for  $\sigma$ , as illustrated in Fig. 1. According to the entropy theory, when the potential entropy is maximum, the uncertainty is also maximum and the network distribution tends to be uniform. In that case, we have  $\frac{\varphi(v_i)}{Z} = \frac{1}{n}$ . Therefore, we will take  $\sigma$  when the potential entropy is minimum in our method (see Fig. 1.). According to the definition of potential entropy, we have:

- 1) When  $\sigma \rightarrow 0^+$ ,  $\varphi(u_i \rightarrow u_j) \rightarrow 0$ , there will be no interaction between nodes  $u_i$  and  $u_j$ , and we have  $\varphi(i) = (m_i)^2 = M^2$ . Thus, the potential entropy will approach the maximum value  $\log(n)$ ;
- 1) When  $\sigma \rightarrow +\infty$ ,  $\varphi(j \rightarrow i) \rightarrow m_j$ , then no matter what the distance between two nodes is, their interaction force will be the same, and we have  $\varphi(i) = nM^2$ . If we normalize  $Z$ , the potential entropy will still approach the maximum value  $\log(n)$ .

Therefore, the potential entropy is a function of  $\sigma$ . The range of  $\sigma$  is  $(0, +\infty)$  and the range of potential entropy is  $(0, \log(n))$ . The value of potential entropy will first decrease monotonically with the increase of  $\sigma$ . However, the value of potential entropy will increase monotonically with the increase of  $\sigma$ , when the minimum value is reached. The potential entropy reaches the maximum value at both ends of  $\sigma$ 's curve.

Thus, we can further identify influential nodes from OSNs according to their topological potential values. The details for predicting influential nodes is illustrated in Algorithm I. With the algorithm, we can finally get the top  $n\%$  items as the recommended influential nodes for brand communication.

#### IV. PERFORMANCE EVALUATION AND INDUSTRIAL APPLICATIONS

In this section, we present various experiments to evaluate the performance of the proposed TPS method on a real-world dataset from SMZDM.com.

##### A. Experimental Setup

To evaluate the proposed TPS, we collected a real dataset from SMZDM.com to carry out the experiments. SMZDM.com is an online shopping guide website in China that also integrates product review services such as Yelp and social network services similar to Facebook and Twitter. We have implemented a crawling program based on Python to crawl brand-related content from the website automatically. The data we extracted are all related to Xiaomi, which is a well-known and typical mobile phone brand in China. We extracted the posts about Xiaomi within a period of time (until August 25, 2019). Thus, we obtained a brand communication dataset for Xiaomi from SMZDM.com to evaluate the performance of our method.

We also processed the original dataset by following the steps illustrated in Section III-A. An open-source Chinese language segmentation tool was used to deal with the posts from the OSN. The number of nodes in the extracted dataset is approximately 40181, and the number of edges is about 60000. Among them, the number of edges with weights

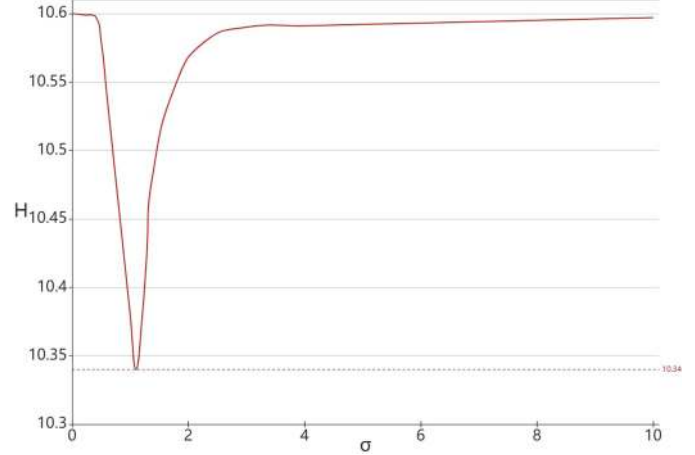


Fig. 1. The relationship between the potential entropy  $H$  and the influence factor  $\sigma$ .

greater than or equal to 2 is approximately 37812, accounting for 63% of the total number of edges in the dataset; in addition, the network density is  $3.72 \times 10^{-5}$ . The noise-reduced dataset has 15895 nodes and 37812 edges in total.

##### B. Network Characteristics Analysis

We first try to divide the dataset into several subcommunities and verify the scale-free and small-world properties of these subcommunities. We used the Gephi software to generate an interaction network diagram for the brand communication dataset, as shown in Fig. 2. There exist many subcommunities in this network. We use the modular function of Gephi to divide the subcommunities. By setting the three parameters of *Randomize*, *Use edge weights* and *Resolution* in the software, we find that the modularity of the network and the modularity with resolution are both 0.757, and the number of subcommunities is 1155 (see Fig. 3). As shown in Fig. 3, the number of nodes in most communities is too small; therefore, we only analyze the eight largest subcommunities. As illustrated in Table I, the sum of the internal degree of each community is much larger than the sum of the external degree.

We further analyze the small-world property of the network for brand communication from an empirical perspective. Table II shows the statistical results of the network statistical properties of the eight largest subcommunities; the maximum value of the average path length in the eight subcommunities is 2.5, which means that one node can reach any other nodes only by 2.5 hops in a subcommunity. We also obtain the clustering coefficients  $C \in (0.008, 0.038)$  for the eight subcommunities. In contrast, the clustering coefficients  $C_{rand}$  of the random networks at the same scale are relatively small. Therefore, it can be concluded that the 8 subcommunities demonstrate the characteristics of a small world, and the information in a subcommunity can be quickly spread to each part of the subcommunity.

The scale-free characteristics of the network are also analyzed through experiments. Fig. 4 and Fig. 5 show the Complementary Cumulative Distribution Function (CCDF) graphs of node indegree and node outdegree, respectively, for the eight subcommunities. By performing a least squares fit on the

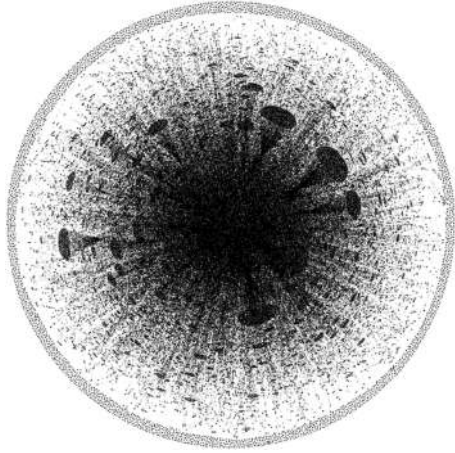


Fig. 2. The interactive network diagram for the brand communication dataset.

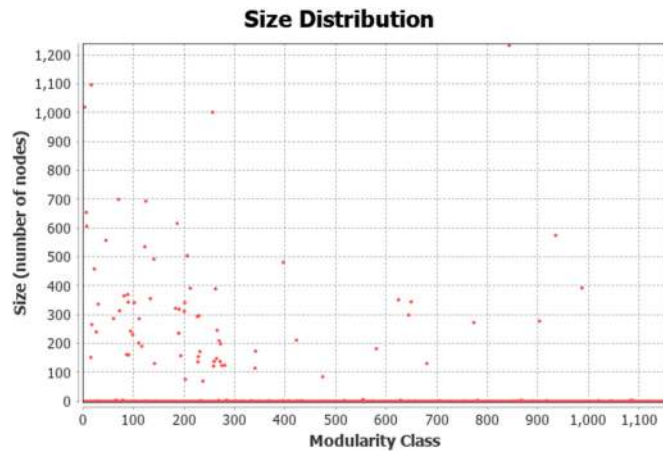


Fig. 3. The results for subcommunity modularity.

node set, we can get the expression for the fitted curve as follows:

$$P[X > x] \approx cx^{-\alpha}. \quad (15)$$

According to 15, we have the power-law exponent  $\alpha > 0$  of the indegree and outdegree distribution for the eight subcommunities (see Table III), which indicates that there are fewer nodes with a larger indegree and more nodes with a smaller indegree, which is consistent with the scale-free feature for social networks. In other words, only a few members have deep participation in the network for brand communication, and they are only the promoters of the development for the brand community. The statistical results show the correlation coefficient  $\gamma < 0$  of the eight subcommunities; that is, the nodes with higher degrees are mostly connected with the nodes with lower degrees. In other words, in the process of information spreading, the information tends to flow from influential nodes to common nodes in the network.

### C. Influential Node Identification

By using the proposed method, we selected the top 20 nodes from the candidate set as the influential nodes, as shown in

TABLE I  
INFORMATION ABOUT SUBCOMMUNITY DIVISION

SUBCOMMUNITY ID	NODE NUMBER	SUM OF INTERNAL DEGREE	SUM OF EXTERNAL DEGREE
841	1090	3434	811
14	1100	3235	1147
1	1023	3159	695
254	1005	2682	984
68	703	1792	374
122	697	1893	695
4	658	1891	763
184	620	1690	586

Table IV. We further divide the top 20 nodes into two groups. The first group of nodes has high individual values. According to 13, the nodes with high individual values are more likely to be identified as influential nodes. For example, it can be seen from Table IV that nodes 9339612697 and 6390492327 have the highest topological potential values among the 20 nodes. Their brand engagement scores are also larger than those of other nodes. This means that they have published many posts related to the Xiaomi brand, which are supported by many other users in the network. The second group of nodes does not have high individual values, and some of them even have a low individual value. After investigating these nodes further, we find that they have published few posts about the brand but have often commented on brand-related content. For example, the brand engagement score of 6195251507 is 0, which means that the user has not published any brand-related content or that the content has not received any positive comments. These kinds of nodes are usually ignored by the existing methods and thus will not be identified as influential nodes. Although these nodes rarely publish brand-related content directly, they are very concerned about the brand, and their comments can also be an important part of brand marketing in OSNs.

We have also identified the top 20 nodes by using two different metrics separately rather than their topological potential values (see Table V). The influential nodes identified by using network structure scores and individual values are quite similar, as there are 14 nodes in common for the first and second columns of Table V. Moreover, we can see that the influential nodes identified by using their topological potential values are quite different from those identified by purely using network structure scores or individual values. The first and third columns have 6 nodes in common, while the second and third columns have 8 nodes in common. It makes sense that a node with a high individual value is more likely to be identified as an influential node. However, it is insufficient to consider the individual value of a single node. The proposed method also considers the individual value of surrounding nodes by using the topological potential model and thus can obtain a more accurate result, compared with using pure individual values.

### D. Performance Evaluation

We also compare the performance of the proposed TPS with three existing methods for measuring node importance, namely, Weighted PageRank [53], Weighted HITS [54] and IMUD [2]. The top 20 influential nodes identified by the four different

TABLE II  
ANALYSIS ON THE NETWORK STATISTICAL PROPERTIES FF SUBCOMMUNITIES

SUBCOMMUNITY ID	AVERAGE PATH LENGTH	AVERAGE WEIGHTED DEGREE	CLUSTERING COEFFICIENT	RANDOM NETWORK CLUSTERING COEFFICIENT	CORRELATION COEFFICIENT
841	1.351	2.778	0.01	0.008	-0.323
14	2.262	2.941	0.038	0.012	-0.3
1	1.445	3.088	0.008	0.002	-0.427
254	1.322	2.669	0.013	0.007	-0.367
68	1.65	2.549	0.029	0.018	-0.548
122	1.961	2.716	0.008	0.002	-0.2
4	2.529	2.874	0.02	0.007	-0.213
184	2.511	2.716	0.014	0.009	-0.272

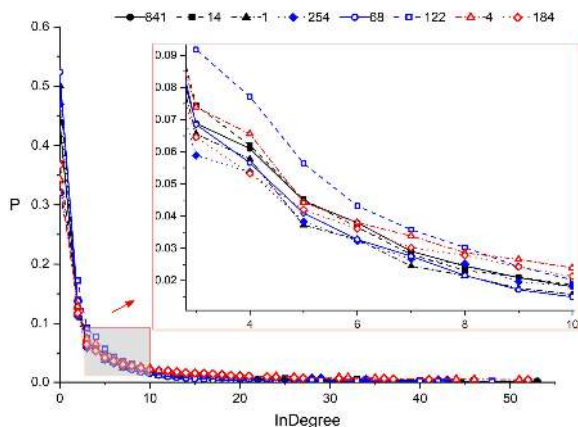


Fig. 4. Complementary distribution function of subcommunity indegree.

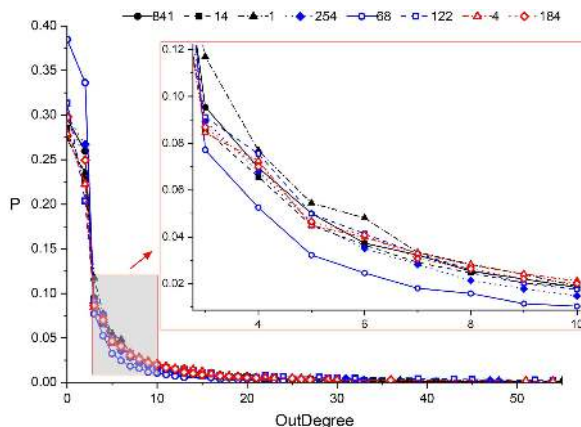


Fig. 5. Complementary distribution function of subcommunity outdegree.

methods are shown in Table VI. The influential nodes identified by Weighted PageRank and Weighted HITS are quite similar, as there are 13 nodes in common for the second and third columns of Table VI. There are also 10 nodes in common for the result sets by Weighted PageRank, Weighted HITS and IMUD. In both result sets by both Weighted PageRank and IMUD, there are only 6 influential nodes in common with the that of TPS. The result set by Weighted HITS has 8 influential nodes in common with that of TPS, which is a little bit larger than that of Weighted PageRank and IMUD.

TABLE III  
POWER-LAW EXPONENT  $\alpha$  OF THE INDEGREE AND OUTDEGREE DISTRIBUTION FOR THE EIGHT SUBCOMMUNITIES

SUBCOMMUNITY ID	$\alpha$ FOR INDEGREE	$\alpha$ FOR OUTDEGREE
1	0.49	0.56
4	0.49	0.51
14	0.54	0.57
68	0.47	0.48
122	0.51	0.52
184	0.48	0.51
254	0.51	0.55
841	0.54	0.58

We have also checked the top 20 influential nodes by the other three methods and find that few of them had published or shared enough content about mobile phone or the Xiaomi brand. For example, the first influential user identified by Weighted PageRank, Weighted HITS and IMUD is the same and it is the user 4077360552, while the first one identified by TPS is 9339612697. Although both users have published many posts about mobile phones, the  $p_{support}$  value of 9339612697 is much larger than that of 4077360552. The posts of 9339612697 receive more positive comments than those of 4077360552. Moreover, 9339612697 has also written more comments on others' posts than 4077360552. Therefore, 9339612697 is more suitable for promoting mobile phone brand like Xiaomi on social media. Moreover, 9339612697 is also the third influential user identified by Weighted PageRank, and the second one by Weighted HITS and IMUD. It also depicts that the influential users identified by TPS are reliable. Both Weighted PageRank and Weighted HITS only address the relationship between nodes, but they do not take into account the content features of users' posts. Therefore, most influential nodes identified by simply using either Weighted PageRank or Weighted HITS are not very valuable for brand communication. Although IMUD has taken into account the content-related features like the topics and the related messages exchanged by users, they have neglected the factors like sentiment and the impact of surrounding nodes.

According to our investigation, there are no widely accepted metrics used to evaluate the performance of influential node mining. In this article, we use the *ratio of*



TABLE IV  
TOP 20 INFLUENTIAL NODES FOR BRAND COMMUNICATION

RANK	USER ID	SUB-COMMUNITY	$score_{network}$	$score_{brand}$	$value_{indv}$	$\Phi(u)$
1	9339612697	254	12484.5	22.95	33.354	2769.278
2	6390492327	4	942.0	20.88	21.665	1319.462
3	6195251507	901	34696.0	0	28.913	834.102
4	9264719054	254	48684.5	4.50	45.070	666.516
5	7878885949	14	31909.5	14.10	40.691	594.482
6	9941585346	184	13832.0	14.40	25.927	554.993
7	4278646865	4	2951.0	3.92	6.379	529.646
8	6649709521	5	170.0	5.52	5.662	470.074
9	3693789970	184	87.5	4.90	4.973	412.889
10	7702845051	191	83.0	7.47	7.539	409.835
11	8695487703	841	239.5	6.44	6.640	404.370
12	9072194639	99	7872.0	12.88	19.440	377.914
13	3435299983	210	19596.5	2.58	18.910	367.619
14	3365325660	86	17579.0	8.10	22.749	348.034
15	1741709098	184	20972.0	6.88	24.357	327.536
16	7331178137	87	14075.5	2.88	14.610	325.422
17	9829574905	70	153.5	9.70	9.828	321.348
18	3675983550	254	2238	1.82	3.685	305.956
19	8899137147	254	3215	0.90	3.579	297.169
20	3139225531	30	5822.5	0	4.852	295.506

TABLE V  
TOP 20 INFLUENTIAL NODES BY USING DIFFERENT METRICS

TOP 20 NODES BY NETWORK STRUCTURE SCORE	TOP 20 NODES BY INDIVIDUAL VALUE	TOP 20 NODES BY TOPOLOGICAL POTENTIAL VALUE
5776984780	5776984780	9339612697
3530184984	4077360552	6390492327
8542595703	3530184984	6195251507
9264719054	8542595703	9264719054
4915662176	9264719054	7878885949
6872473601	7878885949	9941585346
6195251507	4915662176	4278646865
4855429430	6872473601	6649709521
7878885949	9339612697	3693789970
3951496061	7430779363	7702845051
8669208429	6195251507	8695487703
1741709098	4855429430	9072194639
9474218953	9941585346	3435299983
3435299983	8669208429	3365325660
7334373836	3951496061	1741709098
7374065973	1741709098	7331178137
1093739379	3365325660	9829574905
3365325660	4100254403	3675983550
6077130617	7334373836	8899137147
7833474586	6390492327	3139225531

verified users and the ratio of user coverage to evaluate the performance. The ratio of verified users refers to the proportion of verified users among the collection of influential users. The ratio of user coverage refers to the proportion of the users that can be covered or affected by the top  $n\%$  influential nodes among the complete set of users. As seen

from Table VII, the ratio of verified users of TPS is much higher than that of the other three methods. Using the proposed TPS, 1461 out of 2000 influential users are verified. The comparison of the user coverage ratio is illustrated in Fig. 6. The curves of the three methods begin to flatten when  $n \geq 1$ . Therefore, if the top 1% of the influential nodes identified by the four methods are considered separately, the proposed method can directly cover more than 60% of users in the network. However, we can see that TPS can cover more users than the other three methods when  $n \geq 1$ . Additionally, it can be seen that the proposed method can cover almost 100% of users in the sample set when  $n \geq 40$ , while IMUD, Weighted PageRank and Weighted HITS can only cover 89.1%, 86.4%, 86.6% with the same  $n$ . Therefore, we can see that TPS performs better than the other three methods from the perspectives of both the ratio of verified users and the ratio of user coverage.

### E. Industrial Applications

With the popularity of OSNs in our daily life, mining and discovering key opinion leaders or influential nodes from large-scale social networks has become a research hotspot. Currently, increasing number of companies tend to promote their brands or products (especially some newly released ones) through social media instead of traditional media. The method proposed in this article can be applied to support intelligent brand communication or marketing in real-life industrial applications.

Traditionally, when carrying out social media marketing (or viral marketing), companies' marketing information will be pushed to a group of consumers in OSNs. This group of

TABLE VI  
TOP 20 INFLUENTIAL NODES IDENTIFIED BY DIFFERENT METHODS

RANK	TPS	WEIGHTED PAGERANK	WEIGHTED HITS	IMUD
1	9339612697	4077360552	4077360552	4077360552
2	6390492327	5776984780	9339612697	9339612697
3	6195251507	9339612697	7878885949	7878885949
4	9264719054	8542595703	8542595703	8542595703
5	7878885949	7878885949	8982803543	8982803543
6	9941585346	7185586935	3365325660	6907297197
7	4278646865	8982803543	1741709098	5776984780
8	6649709521	9072194639	3990065324	7833474586
9	3693789970	9941585346	5776984780	6460841696
10	7702845051	1741709098	9072194639	1741709098
11	8695487703	6460841696	7833474586	3365325660
12	9072194639	7833474586	6907297197	3990065324
13	3435299983	8253144403	4918744579	4289304618
14	3365325660	3527641579	7185586935	3435299983
15	1741709098	2981714089	3435299983	3951496061
16	7331178137	3365325660	9941585346	9912988565
17	9829574905	8219924777	6489041006	9072194639
18	3675983550	7374065973	6390492327	7358548562
19	8899137147	9381408906	7148994619	4918744579
20	3139225531	3920752057	7930650431	1378211555

TABLE VII  
RATIO OF VERIFIED USERS BY THE THREE METHODS

METHODS	RATIO OF VERIFIED USERS
Weighted PageRank	831/2000
Weighted HITS	599/2000
IMUD	718/2000
TPS	1461/2000

consumers is usually selected by the platform. Although some of them may forward the material to others when they receive marketing information, the effect of information spreading is limited due to their influence on OSNs. Therefore, companies want to choose a set of customers to market to that will maximize their Internet profits (profits from sales minus the costs of marketing).

The metrics and algorithm proposed in this article can be used to mine and identify influential nodes or users in OSNs (see Fig. 7). After mining a collection of influential users from OSNs, these users are considered seed users. What companies need to do next is to establish trust relationships with these influential users and engage them to promote their brands or products spontaneously on social media. In social media or online communities, eWOM generation can be achieved by influential users after they have positive consumption experiences. As an influential user can always affect a number of common consumers, marketing information can spread quickly through social networks. In this way, companies can promote their brands or products with less cost but better effects on social media. Moreover, companies can even perform personalized recommendations on OSNs through influential users.

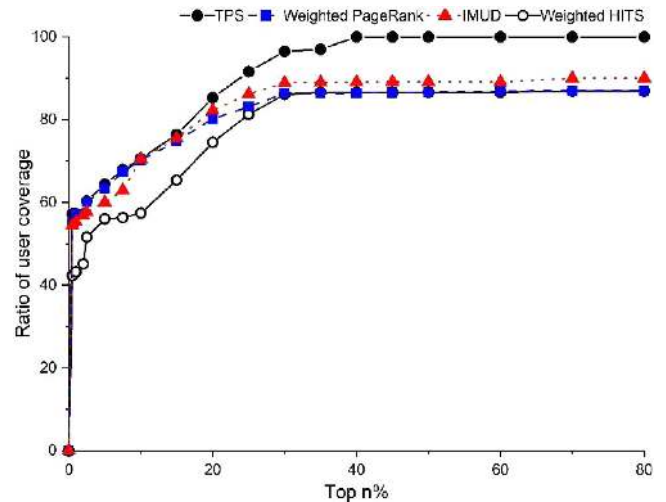


Fig. 6. The comparison of the user coverage ratio for the three methods.

For consumer-oriented industries [55], companies increasingly rely on social media to promote their brands and products. The technology presented in this article is able to mine influential users from large-scale OSNs, and companies can then improve their marketing strategies with the help of those influential users.

## V. CONCLUSION AND FUTURE WORK

In this article, we mainly address the problem of predicting influential nodes from OSNs for brand communication. We quantitatively measure the individual value of nodes by considering both the network structure and content-related factors. Moreover, an improved topological potential scheme is

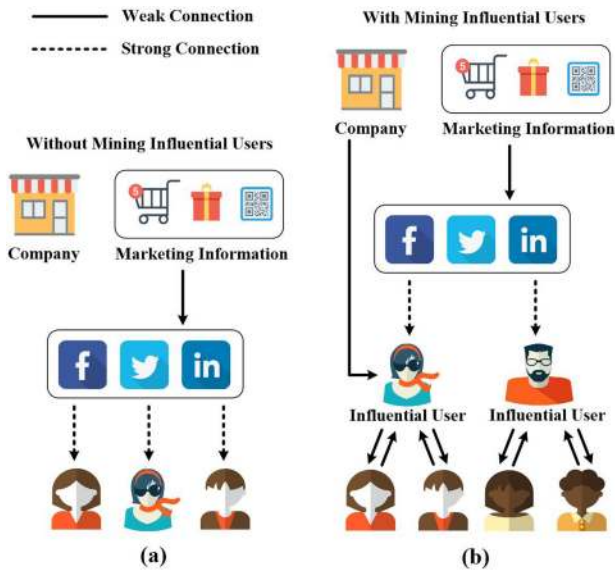


Fig. 7. Application scenario for the proposed method to support more intelligent marketing for companies.

proposed for predicting influential nodes in OSNs. In the process of mining influential nodes from OSNs, network structure, brand engagement, and topological potential are combined together in our method to overcome the limitations of the existing methods. The computational results suggest that the proposed method is able to predict influential nodes for brand communication in OSNs. We can find out the nodes that have published few posts about the brand but commented on brand-related content a lot, which are usually ignored by the existing methods. Moreover, we can obtain identification results that have a higher ratio of verified users and user coverage by using the proposed method compared to three existing methods.

We also consider some possible future directions of this study. For example, we only used the followship and comment relationship between users to model the weighted network. In fact, there exist more deep or potential relationships among users, which can be discovered by using more complex mining algorithms. Therefore, in the future, we can obtain a more complex network model for predicting influential nodes. Additionally, we only investigated the characteristics of users in OSNs statically, but we have not considered the impact of time changes. If we take into account the time factor and study the time-dependent trend of user behaviors in OSNs, we can obtain more characteristic information about influential nodes.

## REFERENCES

- [1] A. Aleahmad *et al.*, "OLFfinder: Finding opinion leaders in online social networks," *J. Inf. Sci.*, vol. 42, no. 5, pp. 659–674, 2016.
- [2] A. Zareie, A. Sheikahmadi, and M. Jalili, "Identification of influential users in social networks based on users' interest," *Inf. Sci.*, vol. 493, pp. 217–231, 2019.
- [3] W. Guo, Y. Shi, S. Wang, and N. N. Xiong, "An unsupervised embedding learning feature representation scheme for network big data analysis," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 115–126, Jan.-Mar. 2020.

- [4] K. Huang, Q. Zhang, C. Zhou, N. Xiong, and Y. Qin, "An efficient intrusion detection approach for visual sensor networks based on traffic pattern learning," *IEEE Trans. Syst., Man, Cybern.: Syst.*, vol. 47, no. 10, pp. 2704–2713, Oct. 2017.
- [5] S. M. H. Bamakan, I. Nurgaliev, and Q. Qu, "Opinion leader detection: A methodological review," *Expert Syst. Appl.*, vol. 115, pp. 200–222, 2019.
- [6] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Netw.*, vol. 1, no. 3, pp. 215–239, 1978.
- [7] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.
- [8] H. W. Corley and D. Y. Sha, "Most vital links and nodes in weighted networks," *Oper. Res. Lett.*, vol. 1, no. 4, pp. 157–160, 1982.
- [9] Y. Zeng *et al.*, "Maximizing influence under influence loss constraint in social networks," *Expert Syst. Appl.*, vol. 55, pp. 255–267, 2016.
- [10] E. Even-Dar and A. Shapira, "A note on maximizing the spread of influence in social networks," *Inf. Process. Lett.*, vol. 111, no. 4, pp. 184–187, 2011.
- [11] J. Zhu, J. Zhu, S. Ghosh, W. Wu, and J. Yuan, "Social influence maximization in hypergraph in social networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 6, no. 4, pp. 801–811, Oct.-Dec. Jan. 2019.
- [12] A. Zareie, A. Sheikahmadi, and M. Jalili, "Influential node ranking in social networks based on neighborhood diversity," *Future Gener. Comput. Syst.*, vol. 94, pp. 120–129, 2019.
- [13] Y. Yang, N. Xiong, N. Y. Chong, and X. Défago, "A decentralized and adaptive flocking algorithm for autonomous mobile robots," in *Proc. 3rd Int. Conf. Grid Pervasive Comput.*, 2008, pp. 262–268.
- [14] S. Kumar and B. S. Panda, "Identifying influential nodes in social networks: Neighborhood coreness based voting approach," *Phys. A: Stat. Mechanics Appl.*, vol. 553, 2020, Art. no. 124215.
- [15] C. Salavaty and A. Abdollahpouri, "Identifying influential nodes based on ant colony optimization to maximize profit in social networks," *Swarm Evol. Comput.*, vol. 51, 2019, Art. no. 100614.
- [16] B. Zhang *et al.*, "A most influential node group discovery method for influence maximization in social networks: A trust-based perspective," *Data Knowl. Eng.*, vol. 121, pp. 71–87, 2019.
- [17] A. Salavaty, M. Ramialison, and P. D. Currie, "Integrated value of influence: An integrative method for the identification of the most influential nodes within networks," *Patterns*, vol. 1, no. 5, 2020, Art. no. 100052.
- [18] F. Zhou, L. Lu, and M. S. Mariani, "Fast influencers in complex networks," *Commun. Nonlinear Sci. Numer. Simul.*, vol. 74, pp. 69–83, 2019.
- [19] I. G. Amnieh and M. Kaedi, "Using estimated personality of social network members for finding influential nodes in viral marketing," *Cybern. Syst.: Int. J.*, vol. 46, no. 5, pp. 355–378, 2015.
- [20] Q. Zhang, Q. Qiu, W. Guo, K. Guo, and N. Xiong, "A social community detection algorithm based on parallel grey label propagation," *Comput. Netw.*, vol. 107, Part 1, pp. 133–143, 2016.
- [21] J. Lloret *et al.*, "Improving networks using group-based topologies," *Comput. Commun.*, vol. 31, no. 14, pp. 3438–3450, 2008.
- [22] L. Jain and R. Katarya, "Discover opinion leader in online social network using firefly algorithm," *Expert Syst. Appl.*, vol. 122, pp. 1–15, 2019.
- [23] S. Srinivas and C. Rajendran, "Community detection and influential node identification in complex networks using mathematical programming," *Expert Syst. Appl.*, vol. 135, pp. 296–312, 2019.
- [24] Y. Zhao, S. Li, and F. Jin, "Identification of influential nodes in social networks with community structure based on label propagation," *Neurocomputing*, vol. 210, pp. 34–44, 2016.
- [25] Y. Liu, A. Liu, N. Xiong, T. Wang, and W. Gui, "Content propagation for content-centric networking systems from location-based social networks," *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 49, no. 10, pp. 1946–1960, Oct. 2019.
- [26] S. Liu *et al.*, "Identifying effective influencers based on trust for electronic word-of-mouth marketing: A domain-aware approach," *Inf. Sci.*, vol. 306, pp. 34–52, 2015.
- [27] R. Narayanam and A. A. Nanavati, "Design of viral marketing strategies for product cross-sell through social networks," *Knowl. Inf. Syst.*, vol. 39, no. 3, pp. 609–641, 2014.
- [28] H. T. Nguyen, T. N. Dinh, and M. T. Thai, "Cost-aware targeted viral marketing in billion-scale networks," in *Proc. 35th Annu. IEEE Int. Conf. Comput. Commun.*, 2016, pp. 1–9.
- [29] J. Weng *et al.*, "Twitterrank: Finding topic-sensitive influential twitterers," in *Proc. 3rd Int. Conf. Web Search Web Data Mining*, 2010, pp. 261–270.
- [30] X. Song *et al.*, "Identifying opinion leaders in the blogosphere," in *Proc. 16th ACM Conf. Conf. Inf. Knowl. Manage.*, 2007, pp. 971–974.

- [31] H. Zhou, D. Zeng, and C. Zhang, "Finding leaders from opinion networks," in *Proc. IEEE Int. Conf. Intell. Secur. Inform.*, 2009, pp. 266–268.
- [32] B. Huang, G. Yu, and H. R. Karimi, "The finding and dynamic detection of opinion leaders in social network," *Math. Problems Eng.*, vol. 2014, 2014, Art. no. 328407.
- [33] C. Wang, Y. J. Du, and M. W. Tang, "Opinion leader mining algorithm in microblog platform based on topic similarity," in *Proc. 2nd IEEE Int. Conf. Comput. Commun.*, 2016, pp. 160–165.
- [34] N. Ma and Y. Liu, "SuperedgeRank algorithm and its application in identifying opinion leader of online public opinion supernetwork," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1357–1368, 2014.
- [35] Y. Li *et al.*, "An improved mix framework for opinion leader identification in online learning communities," *Knowl.-Based Syst.*, vol. 43, pp. 43–51, 2013.
- [36] A. Hamzehei, S. Jiang, D. Koutra, R. Wong, and F. Chen, "Topic-based social influence measurement for social networks," *Australas. J. Inf. Syst.*, vol. 21, 2017, doi: [10.3127/ajis.v21i0.1552](https://doi.org/10.3127/ajis.v21i0.1552).
- [37] Q. Fang *et al.*, "Topic-sensitive influencer mining in interest-based social media networks via hypergraph learning," *IEEE Trans. Multimedia*, vol. 16, no. 3, pp. 796–812, Apr. 2014.
- [38] A. Shahzad *et al.*, "Real time MODBUS transmissions and cryptography security designs and enhancements of protocol sensitive information," *Symmetry*, vol. 7, no. 3, pp. 1176–1210, 2015.
- [39] A. Hajikhani, J. Porras, and H. Melkas, "Brand analysis in social network services: Results from content analysis in twitter regarding the us smartphone market," *Int. J. Innov. Technol. Manage.*, vol. 14, no. 2, 2017, Art. no. 1740008.
- [40] S. Kabadayi, and K. Price, "Consumer-brand engagement on facebook: Liking and commenting behaviors," *J. Res. Interactive Marketing*, vol. 8, no. 3, pp. 203–223, 2014.
- [41] B. Schivinski and D. Dabrowski, "The effect of social media communication on consumer perceptions of brands," *J. Marketing Commun.*, vol. 22, no. 2, pp. 189–214, 2016.
- [42] D. Jiménez-Castillo and R. Sánchez-Fernández, "The role of digital influencers in brand recommendation: Examining their impact on engagement, expected value and purchase intention," *Int. J. Inf. Manage.*, vol. 49, pp. 366–376, 2019.
- [43] Q. Gao and C. Feng, "Branding with social media: User gratifications, usage patterns, and brand message content strategies," *Comput. Human Behav.*, vol. 63, pp. 868–890, 2016.
- [44] B. Godey *et al.*, "Social media marketing efforts of luxury brands: Influence on brand equity and consumer behavior," *J. Bus. Res.*, vol. 69, no. 12, pp. 5833–5841, 2016.
- [45] M. De Veirman, V. Cauberghe, and L. Hudders, "Marketing through instagram influencers: The impact of number of followers and product divergence on brand attitude," *Int. J. Advertising*, vol. 36, no. 5, pp. 798–828, 2017.
- [46] J. Jacoby and R. W. Chesnut, "Brand loyalty: Measurement and management," New York, USA: John Wiley & Sons, 1978.
- [47] J. van Doorn *et al.*, "Customer engagement behavior: Theoretical foundations and research directions," *J. Service Res.*, vol. 13, no. 3, pp. 252–266, 2010.
- [48] D. Xue, S. Hirche, and M. Cao, "Opinion behavior analysis in social networks under the influence of cooperative media," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 3, pp. 961–974, Jul./Sep. Jan. 2020.
- [49] W. Wu, N. Xiong, and C. Wu, "Improved clustering algorithm based on energy consumption in wireless sensor networks," *IET Netw.*, vol. 6, no. 3, pp. 47–53, 2017.
- [50] D. Lin, B. Ma, M. Jiang, N. Xiong, K. Lin, and D. Cao, "Social network rumor diffusion predication based on equal responsibility game model," *IEEE Access*, vol. 7, pp. 4478–4486, 2019.
- [51] Y. Wang, A. V. Vasilakos, J. Ma, and N. Xiong, "On studying the impact of uncertainty on behavior diffusion in social networks," *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 45, no. 2, pp. 185–197, Feb. 2015.
- [52] Y. Mao, S. Bolouki, and E. Akyol, "Spread of information with confirmation bias in cyber-social networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 2, pp. 688–700, Apr.–Jun. Jan. 2020.
- [53] W. Xing and A. A. Ghorbani, "Weighted pagerank algorithm," in *Proc. 2nd Annu. Conf. Commun. Netw. Serv. Res.*, 2004, pp. 305–314.
- [54] D. Tsuyoshi *et al.*, "Hubs and authorities in the world trade network using a weighted HITS algorithm," *PLoS One*, vol. 9, no. 7, 2014, Art. no. e100338.
- [55] Q. Zhang, C. Zhou, N. Xiong, Y. Qin, X. Li, and S. Huang, "Multimodel-based incident prediction and risk assessment in dynamic cybersecurity protection for industrial control systems," *IEEE Trans. Syst., Man, Cybern.: Syst.*, vol. 46, no. 10, pp. 1429–1444, Oct. 2016.



**Yuxin Mao** received the B.S. and Ph.D. degrees in computer science from Zhejiang University, Hangzhou, China, in 2003 and 2008, respectively. He is an Associate Professor with the School of Management and E-Business, Zhejiang Gongshang University, Hangzhou, China. His research interests include social network analysis, text mining and Internet marketing. He is a Senior Member of China Computer Federation and the Member of China Society of Logistics.



**Lujie Zhou** received the B.S. degree in information management from Wenzhou Medical University, Wenzhou, China, in 2015 and the M.S. degree in management science and engineering from Zhejiang Gongshang University, Hangzhou, China, in 2019. Her research interests include data mining, social network analysis, and e-commerce.



**Naixue Xiong** received both the Ph.D. degrees in sensor system engineering from Wuhan University in 2007, and Japan Advanced Institute of Science and Technology in dependable communication networks in 2008, respectively. He is currently an Associate Professor (5rd year) with the Department of Mathematics and Computer Science, Northeastern State University, OK, USA. Before he attended Northeastern State University, he worked with Georgia State University, Wentworth Technology Institution, and Colorado Technical University (full professor about 5 years) about 10 years. He has authored or coauthored more than 200 international journal papers and more than 100 international conference papers. His research interests include cloud computing, security and dependability, parallel and distributed computing, networks, and optimization theory. Some of his works were published in IEEE JSAC, IEEE or ACM transactions, ACM Sigcomm workshop, IEEE INFOCOM, ICDCS, and IPDPS. He has been a General Chair, Program Chair, Publicity Chair, Program Committee member and Organizing Committee Member of more than 100 international conferences, and as a Reviewer of about 100 international journals, including IEEE JSAC, IEEE SMC (Park: A/B/C), IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE TRANSACTION ON PARALLEL AND DISTRIBUTED SYSTEMS. He is an Editor-in-Chief, Associate Editor or Editor Member for more than 10 international journals (including Associate Editor for IEEE TRANSACTION ON SYSTEMS, MAN & CYBERNETICS: SYSTEMS, Associate Editor for *Information Science*, Editor-in-Chief for *Journal of Internet Technology (JIT)*, and Editor-in-Chief for *Journal of Parallel & Cloud Computing (PCC)*), and a Guest Editor for more than 10 international journals, including Sensor Journal, WINET and MONET. He was the recipient of the the Best Paper Award in the 10th IEEE International Conference on High Performance Computing and Communications (HPCC-08) and the Best student Paper Award in the 28th North American Fuzzy Information Processing Society Annual Conference (NAFIPS2009).